LINKAGE ANALYSIS USING AFFECTED SIB-PAIRS

Pak Sham, Jinghua Zhao
Institute of Psychiatry, London.

1 Introduction

In recent years, affected sib-pairs have become increasingly used for the linkage analysis of common diseases such as diabetes and asthma, with the purpose of localizing and characterizing the genetic components of these complex phenotypes. This increasing use of affected sib-pairs has been accompanied by the development of statistical methods and computer software for data analysis. This approach has already yielded a number of promising positive linkage findings, some of which are likely to be replicated and to lead to advances in our understanding of the aetiology of these disorders.

This chapter is an introductory account of non-parametric methods of linkage analysis for affected sib-pairs. First, the key concept of genetic identity-by-descent (IBD) will be explained. Next, some currently popular analytical methods and computer programs will be described. Finally, the strengths and limitations of the approach in general will be discussed. The reader is assumed to have an understanding of the principles of linkage analysis using the standard parametric (i.e. lod score) method.

2 Genetic identity-by-descent between sib-pairs

Non-parametric linkage analysis is based on the detection of an association between the sharing of disease status and the sharing of marker alleles by relatives. There are two different measures of allele-sharing, identity-by-state (IBS) or identity-by-descent (IBD). Two alleles of the same form (i.e. having the same DNA sequence) are said to be IBS. If, in addition to being IBS, two alleles are descended from (and are therefore replicates of) the same ancestral allele, then they are said to be IBD. The general idea of non-parametric linkage analysis is that, in the vicinity of a disease locus, sib-pairs who are concordant for disease status (i.e. both affected or both unaffected) should show an increase in allele-sharing, and those who are discordant for disease status (i.e. one affected and one unaffected) should show a decrease in allele-sharing, from the level of allele-sharing expected of sib-pairs. Of the two measures of allele-sharing, IBS is easier to define and determine, but IBD is more directly relevant to linkage.

In the ideal situation where the marker is extremely polymorphic so that all the alleles in the founding members of a pedigree are distinguishable from each other, IBS implies IBD, and the distinction between the two concepts becomes unnecessary. Consider two parents with genotypes S_1S_2 and S_3S_4. Mendel's law of segregation implies that each offspring of these two parents is equally likely to have the 4 genotypes S_1S_3, S_1S_4, S_2S_3, S_2S_4. The possible combinations of genotypes for a sib-pair, and the corresponding number of alleles IBS (and hence IBD) are as follows

Sibling 1	Sibling 2			
	S_1S_3	S_1S_4	S_2S_3	S_2S_4
S_1S_3	2	1	1	0
S_1S_4	1	2	0	1
S_2S_3	1	0	2	1
S_2S_4	0	1	1	2

Each of these combinations has a probability of (1/4)(1/4)=1/16, so that the probability of the sib-pair having 0, 1 and 2 alleles IBD are 4/16, 8/16 and 4/16, respectively. This 1/4, 1/2, 1/4 distribution for the IBD values of 0, 1, 2 represents the expected level of IBD for a sib-pair from which any excess or deficit in allele-sharing must be measured.

In reality, DNA markers are never sufficiently polymorphic to ensure that all founder alleles are distinguishable from each other. This has the undesirable effect that the true IBD value (whether 0, 1 or 2) of a sib-pair cannot always be determined from the marker genotypes of the family. In this more realistic scenario, it becomes important to distinguish IBS from IBD, as the former no longer implies the latter. The distinction between IBS and IBD can be illustrated by the following nuclear family (Figure 1).

The two offspring of this family are IBS for allele S_1. The parental origins of these two S_1 alleles are, however, different. While the S_1 allele of the first offspring (genotype S_1S_2) is derived from the maternal S_1 allele, the S_1 allele of the second offspring (with genotype S_1S_3) is derived from the paternal S_1 allele. On the assumption that the two parents are indeed unrelated and non-inbred, these two S_1 alleles must be considered non-IBD. The sib-pair in this family is therefore IBS for 1 allele (S_1), but IBD for 0 allele.

In order to determine the IBS status of a sib-pair, it is sufficient to have genotypic data on the sib-pair alone. This is not true for IBD status. Suppose the parental genotypes for the illustrative family in Figure 1 are (S_1S_4, S_2S_3) instead of (S_1S_2, S_1S_3). In this case the two offspring must have

inherited the same allele (S_1) from the father, so that the sib-pair is IBD for 1 allele. The genotypic data on the sib-pair (i.e. S_1S_2, S_1S_3) are therefore consistent with an IBD status of 0 and 1, depending on the genotypes of the parents. In general, if the IBS value of a sib-pair is 0, then the IBD value the sib-pair must also be 0, but if the IBS value of a sib-pair is 1 or 2, then the IBD value of the sib-pair cannot be determined without the parental genotypes.

3 Genetic identity by descent and recombination fraction

Under the assumption that the two parents of a sib-pair are unrelated and non-inbred, the four parental alleles can be considered as four distinct ancestral alleles (whether or not they are IBS). The sib-pair can be IBD for 0, 1, or 2 alleles at each chromosomal location in the genome. The usefulness of IBD values for linkage analysis is based on the fact that the IBD values (0, 1 or 2) of a sib-pair are independent for non-syntenic loci, but are positively correlated for syntenic loci. The closer the two loci, the greater is the correlation between their IBD values. It can be shown that, when the recombination fraction between loci A and B is θ, the conditional distribution of IBD values at B (denoted as D_B), given the IBD value at A (denoted as D_A), is as follows (Table 1).

Table 1: Conditional IBD distribution for sib-pair

	$P(D_B\|D_A)$		
	$D_B = 0$	$D_B = 1$	$D_B = 2$
$D_A = 0$	Ψ^2	$2\Psi(1-\Psi)$	$(1-\Psi)^2$
$D_A = 1$	$\Psi(1-\Psi)$	$1-2\Psi(1-\Psi)$	$\Psi(1-\Psi)$
$D_A = 2$	$(1-\Psi)^2$	$2\Psi(1-\Psi)$	Ψ^2

where Ψ is defined as $\theta^2 + (1-\theta)^2$. From this it can be shown that the correlation between D_A and D_B is $2\Psi - 1$. This correlation ranges from the value of 0 when $\theta = 1/2$, to the value of 1 when $\theta = 0$. The positive correlation between the IBD values of two tightly linked loci has the important implication that an affected sib-pair is expected to show an increase in IBD not only at the disease locus itself but also at neighbouring markers.

4 Identity by descent distribution for an affected sib-pair

The most popular family unit for non-parametric linkage analysis is an affected sib-pair, with or without parents. Affected sib-pairs are often far more informative for linkage than unaffected sib-pairs, or sib-pairs with one affected and one unaffected member, under many plausible models of complex inheritance. Intuitively, this is because the presence of disease is often more indicative of the presence of the disease gene than the absence of the disease is indicative of the absence of the disease gene. Moreover, if the disease gene is rare, then it is absent in most parents whose offspring are all unaffected. Most families with two unaffected offspring are therefore uninformative for linkage.

A more precise analysis of the IBD distribution of an affected sib-pair is desirable for two purposes. The first is that the IBD distribution may provide insights that are helpful in the construction of statistical tests for linkage. The second is that it can be used to calculate expected IBD values that are necessary for power analysis and sample size determination. The greater the deviation of the expected IBD distribution from the null hypothesis values of 1/4, 1/2, 1/4, the greater is the power for detecting linkage.

The IBD distribution of an affected sib-pair at a marker locus linked to a disease locus was derived by Suarez et al (1978), under the assumptions of a single-locus disease model and random mating. The parameters of the model are the disease allele frequency (p), the penetrances (f_0, f_1 and f_2, defined as the conditional probabilities of the disease given 0, 1 and 2 copies of the disease allele), and the recombination fraction (θ). The following quantities are then defined in terms of these parameters

$q = 1 - p$
$K = p^2 f^2 + 2pqf_1 + q^2 f_0$
$V_A = 2pq[p(f_2 - f_1) + q(f1 - f_0)]^2$
$V_D = p^2 q^2 (f_2 - 2f_1 + f_0)^2$
$\Psi = \theta^2 + (1 - \theta)^2$

where K is the population prevalence of the disease and V_A and V_D are additive and dominance variance, respectively. Suarez et al (1978) showed that the IBD probabilities for an affected sib-pair are related to these quantities

as follows

$$P(IBD=0) = \frac{1}{4} - \frac{(\Psi - 0.5)V_A + (2\Psi - \Psi^2 - 0.75)V_D}{4(K^2 + 0.5V_A + 0.25V_D)}$$

$$P(IBD=1) = \frac{1}{2} - \frac{2(\Psi^2 - \Psi + 0.25)V_D}{4(K^2 + 0.5V_A + 0.25V_D)}$$

$$P(IBD=2) = \frac{1}{4} + \frac{(\Psi - 0.5)V_A + (\Psi^2 - 0.25)V_D}{4(K^2 + 0.5V_A + 0.25V_D)}$$

Holmans (1993) and Faraway (1993) showed that only limited combinations of IBD probabilities are compatible with these equations. If the IBD probabilities are denoted as z_0, z_1, z_2, then the combinations of IBD values within the triangular region defined by $z_1 \leq 1/2$, $z_1+z_2 \leq 1$ and $(3z_1/2+z_2) \geq 1$ (Figure 2) are compatible, whereas those outside the triangle are not compatible, with a single locus disease model. If it is assumed further that V_D (dominance variance) is zero, then the only admissible combination of IBD values are on the line $z_1 = 1/2$, for $1/4 \leq z_2 \leq 1/2$.

The IBD distribution at the disease locus (i.e. $\theta = 0$) can be re-expressed in terms of the recurrence risks of the disease in various classes of relatives (e.g. monozygotic twins, siblings, offspring) of affected probands (Risch, 1990a). If the recurrence risk of the disease in a relative of class R is denoted as K_R, and the ratio of this risk to the population risk (K) is denoted as λ_R, where the subscript R may take the values M (monozygotic twin), S (sibling), O (offspring), then the IBD probabilities of an affected sib-pair can be re-written in terms of the relative risks $\lambda_M, \lambda_S, \lambda_O$ as follows

$$P(IBD=0) = \frac{1}{4}\frac{1}{\lambda_S}$$

$$P(IBD=1) = \frac{1}{2}\frac{\lambda_O}{\lambda_S}$$

$$P(IBD=2) = \frac{1}{4}\frac{\lambda_M}{\lambda_S}$$

Since these probabilities must sum to 1, it follows that (under a single locus disease model and random mating)

$$\lambda_M = 4\lambda_S - 2\lambda_O - 1$$

The extension of this formulation to multi-locus disease models is simple, especially in the case where the effects of the alleles on risk are multiplicative.

In this case, each locus contributes a "factor" to the relative risk of each class of relative

$$\lambda_R = \lambda_{1R}\lambda_{2R}\lambda_{3R}\ldots$$

The IBD distribution at the $i'th$ disease locus is then

$$P(IBD_i = 0) = \frac{1}{4}\frac{1}{\lambda_{iS}}$$

$$P(IBD_i = 1) = \frac{1}{2}\frac{\lambda_{iO}}{\lambda_{iS}}$$

$$P(IBD_i = 2) = \frac{1}{4}\frac{\lambda_{iM}}{\lambda_{iS}}$$

with the implication that

$$\lambda_{iM} = 4\lambda_{iS} - 2\lambda_{iO} - 1$$

for each constituent locus. The calculation of the IBD distribution at a particular disease locus requires certain assumptions to be made about the "relative risks" (i.e. $\lambda_{iM}, \lambda_{iS}, \lambda_{iO}$) associated with that locus.

Example. Suppose that the empirical relative risks of a disease in monozygotic cotwins, siblings and offspring of affected probands are 45, 10 and 10, and that the "relative risks" of all contributory loci are identical. Under these assumptions, the number of loci, k, must satisfy the relationship $\sqrt[k]{45} = 4\sqrt[k]{10} - 2\sqrt[k]{10} - 1$. The solution of this equation yields $k = 4.4$, when $\lambda_{iM}, \lambda_{iS}$ and λ_{iO} are 2.375, 1.688 and 1.688 for each constituent locus. Under these "relative risks", affected sib-pairs can be shown to have a probability distribution of z_0=.148, z_1=.5 and z_2=.352 for sharing 0, 1 and 2 alleles IBD, at each constituent locus. If a simple test of proportion of z_2=.25 against $z_2 >$.25 is adopted, then the expected value of z_2 (.352) would indicate that a sample size of 420 affected sib-pairs is required for 80% power to detect linkage at a significance level of .0001.

5 Methods of affected sib-pair analysis

There are several variations of the affected sib-pair method, which can be broadly classified into those based on (or closely related to) Pearson chi-squared statistics (counting the numbers of sib-pairs sharing 0, 1, and 2 alleles IBD and comparing these numbers to their expectations) and those based on likelihood ratio statistics (treating IBD probabilities as unknown parameters).

5.1 Simple IBD counting methods

In early versions of the counting methods, families are retained only if the parental genotypes always allow the IBD status of the sib-pair to be fully determined. The necessary conditions for this are that both parents must be heterozygous and that the two parents must not have the same genotype. Let the total number of affected sib-pairs in such families be n, and numbers of pairs with IBD values of 0, 1 and 2 be $n_0, n_1,$ and $n_2 (n = n_0 + n_1 + n_2)$. Several simple test statistics have been proposed to assess whether these numbers deviate from the expected proportions of (1/4, 1/2, 1/4). The simplest of these is the ordinary Pearson chi-squared statistic,

$$S_1 = \sum_i \frac{(n_i - e_i)^2}{e_i}$$

where $e_0, e_1,$ and e_2 are n/4, n/2 and n/4, respectively. This statistic can be referred to a chi-squared distribution with 2 degrees of freedom for a test of linkage.

Another family of test statistics are linear combinations of the form $vn_1 + n_2$ (Knapp et al, 1994). Among these statistics, the most popular is the so-called "mean test", which is obtained by setting $v = 1/2$. Since the mean and variance of $(1/2)n_1 + n_2$ under the null hypothesis are n/2 and n/8, respectively, the test statistic

$$S_2 = \frac{(\frac{1}{2}n_1 + n_2) - \frac{n}{2}}{\sqrt{\frac{n}{8}}}$$

has asymptotically a standard normal distribution under the null hypothesis of non-linkage. This is simply a test of whether the proportion of alleles IBD is 1/2. Since the alternative hypothesis of linkage implies $[(1/2)n_1 + n_2] - n/2 > 0$, the test is one-tailed. This test can be shown to be locally most powerful (i.e. for θ close to 1/2) under all possible single locus models of inheritance, and to be uniformly most powerful for all value of θ when $f_1^2 = f_0 f_2$ (Knapp et al, 1994). In simulation studies, the mean test appears to perform adequately under a wide range of conditions (Suarez & Van Eerdewegh, 1984; Blackwelder & Elston, 1985).

5.2 Partially informative families

The restriction of affected sib-pair analysis to "fully informative" families can result in some loss of information. In some families, for example, those

with parental genotypes $A_1A_2 \times A_3A_3$, IBD status can be deduced for the alleles transmitted from the A_1A_2 parent but not the A_3A_3 parent. In other families, the genotype of an individual (for example, a parent) may be unavailable, but the available genotypes in other family members may allow the probability distribution of the missing genotype to be calculated. An extension of the classical affected sib-pairs method to make use of such partial information was developed by Sandkujl (1989), and implemented in the ESPA program. Another program that uses a similar approach is SIBPAIR in the ANALYZE package written by Terwilliger.

For each affected sib-pair, the ESPA program attempts a probabilistic reconstruction of missing parental or sibling genotypes from the known genotypes of the family members. In practice, this involves calculating the probabilities of the possible genotype configurations, given the known genotypes in the pedigree, using the MLINK program. For each possible configuration, the number of alleles IBD, the number of alleles non-IBD, and the number of alleles whose IBD status is unknown, are counted. We denote these three counts, for configuration i, by n_{i1}, n_{i0} and $(2 - n_{i1} - n_{i0})$. The "estimated" counts of alleles IBD and alleles non-IBD in the sib-pair is then the weighted sums $\sum_i p_i n_{i1}$, and $\sum_i p_i n_{i0}$, where p_i is the probability of configuration i. These counts are then aggregated across all the affected sib-pairs in the sample, to obtain the total "estimated" counts of alleles IBD and alleles non-IBD. If these counts are denoted as N_1 and N_0, then ESPA computes the statistic $(N_1 - N_0)^2/(N_1 + N_0)$ and interprets it as a one-tailed chi-squared test for the hypothesis of linkage.

The ESPA program has many attractive features. It uses standard LINKAGE format pedigree and parameter input files. The final counts N_1 and N_0, and the test statistic $(N_1 - N_0)^2/(N_1 + N_0)$ are easy to interpret, and the program handles pedigree data with multiple affected sib-pairs and missing genotypes. However, it is unclear whether the statistic $(N_1 - N_0)^2/(N_1 + N_0)$ is asymptotically chi-squared, as the counts N_1 and N_0 are "estimates" rather than observations. Moreover, multiple affected sib-pairs in the same family, which may not constitute independent observations, are included in the calculation of N_1 and N_0. This dependence may distort the distribution of the test statistic. Also, the reconstruction of missing parental genotypes from the observed genotypes of the affected sib-pair may cause the test to be biased in favour of the null hypothesis. This is because the reconstruction of parental genotypes is easier for sib-pairs who are non-IBD for both alleles, than for sib-pairs who are IBD for both alleles. Thus, an affected sib-pair with 4 distinct alleles (e.g. A_1A_2, A_3A_4) will "force" the two parents to have these 4 alleles even if parental genotypes

are unavailable, so that the sib-pair will be given full weight in the analysis. On the other hand, an affected sib-pair who are both homozygous with the same allele (e.g. A_1A_1, A_1A_1) may be IBD for both alleles, but whether this is so will be uncertain unless the parental genotypes are available. The weights given to such sib-pairs will therefore be incomplete. The net effect of this bias is a spurious excess of non-IBD alleles. This problem can be avoided by ignoring the genotypes of the affected sib-pair when reconstructing the genotypes of the parents. However, this may result in some loss of information, which may be even more detrimental than the bias in favour of the null hypothesis, especially if the marker is highly polymorphic.

5.3 Likelihood methods

Likelihood methods of affected sib-pair analysis were introduced to provide a more satisfactory way of dealing with the problem of incomplete data (Risch, 1990b; Holmans, 1993). The statistical model is that each affected sib-pair has probabilities z_0, z_1 and z_2 of having 0, 1 and 2 alleles IBD. These probabilities, written as the vector $\mathbf{z}$, are defined as the parameters of the model. The likelihood of a set of genotypic data $\mathbf{x}$ can be written as a function of $\mathbf{z}$ (and the allele frequencies at the marker locus, which are either known or estimated from the data). For a family containing one affected sib-pair, the likelihood can be decomposed as follows

$$L(\mathbf{z}) = \sum_{i=0}^{2} z_i P(\mathbf{x}|IBD = i)$$

which is a linear combination of z_0, z_1 and z_2. If the log-likelihood maximised with respect to $\mathbf{z}$ is denoted as $\ln L_1$ and the log-likelihood at the null hypothesis (z_0=.25, z_1=.5, z_2=.25) is denoted as $\ln L_0$, then the likelihood ratio statistic $2(\ln L_1 - \ln L_0)$ is asymptotically chi-squared with two degrees of freedom and provides a test for linkage. There are only two free parameters because of the obvious constraint $z_0+z_1+z_2=1$.

Two levels of restrictions are often imposed on $\mathbf{z}$. Holmans (1993) suggested constraining the parameters z_1 and z_2 to the triangular region compatible with a single locus disease model (Figure 2). Let the maximum log-likelihood within this region be $\ln L_{1^\star}$, then a likelihood ratio test statistic for linkage is $2(\ln L_{1^\star} - \ln L_0)$. Using the method of Self and Liang (1987), the asymptotic distribution of this statistic was shown to be a mixture of chi-squared distributions with 0, 1 and 2 degrees of freedom (a chi-squared with 0 degree of freedom is defined as 0). The mixing proportions depend on marker allele frequencies, but are approximately .41, .50, .09.

The second level of restriction on z is obtained by the assumption of zero dominance variance, which constrains z_1 to be 1/2, and z_2 to between 1/4 and 1/2. Let the maximum log-likelihood on this line be $\ln L_{1^{\star\star}}$, then a likelihood ratio test statistic for linkage is $2(\ln L_{1^{\star\star}} - \ln L_0)$. The asymptotic distribution of this statistic is a 50:50 mixture of chi-squared distributions with 0 and 1 degrees of freedom.

Programs that have implemented the likelihood method include SPLINK and MAPMAKER/SIBS, with the latter being able to deal with multipoint data (Kruglyak & Lander, 1995). In multipoint analysis, the likelihood function at a chromosomal location is the probability of the genotypic data of the entire marker set in the region over the IBD values at that chromosomal location. The ability of MAPMAKER/SIBS to handle multi-point data enables full IBD information to be extracted from a region, making it the program of choice.

MAPMAKER/SIBS uses input pedigree and parameter files in standard LINKAGE format, and can be set to either of the two levels of restriction on **z**. It can be used to compute a series of likelihood ratio statistics across a chromosomal region, in order to detect areas with evidence of excess allele-sharing. It can also be used to perform exclusion mapping given a set of hypothetical values for **z**.

6 Affected sib-pair analysis: strengths and limitations

The linkage information content of affected sib-pairs, as compared to other types of affected relative pairs, and to pedigrees with a greater number of affected members, is an important issue for the design of linkage studies. Generally speaking, for a gene of major effect (i.e. high penetrance, high "relative risk"), affected sib-pairs contain far less linkage information than more distantly related affected relative pairs or large pedigrees with multiple affected members. As the relative risk decreases, however, affected sib-pairs becomes progressively more attractive, as aetiological heterogeneity becomes more likely for distantly related affected relative pairs, and homozygosity at the disease locus becomes more likely in densely affected pedigrees. Risch (1990c) showed that, at a "sibling relative risk" of 2 or less, affected sib-pairs are equally informative for linkage as other more distantly related affected relative pairs.

The increasing recognition that common genes of minor or modest effects may be involved in common diseases such as diabetes and depression has

led to the increasing use of the affected sib-pair approach. However, it does not appear sensible to adhere strictly to affected sib-pairs, to the exclusion of other potentially informative family units. In other words, it would be wasteful not to include other types of affected relative pairs, or families with three or more affected members, if these were encountered in the process of identifying affected sib-pairs. Since the affected sib- pair methods described above are designed to deal with pedigrees with just two affected siblings, the inclusion of other family types presents new analytical problems that are still currently under investigation. One possible approach is the generalization of statistics based on excess allele-sharing to general pedigrees (Curtis & Sham, 1994; Davis et al, 1996; Kruglyak et al, 1996). Another possible approach is to modify the lod score method by treating allele frequencies and penetrances as nuisance parameters (Curtis & Sham, 1995). At present, the most popular test is probably that based on the NPL_{all} statistic of the GENEHUNTER program. This statistic is a standardised measure of excessive IBD allele-sharing among the affected members in pedigrees.

Another limitation of the affected sib-pair method is that it does not make full use of information from unaffected family members. When the penetrance of the disease allele is low, many unaffected individuals will be non-penetrant carriers of the disease allele, so that the loss of linkage information is expected to be small. If, however, there is a quantitative trait which is affected by the presence or absence of the disease allele, then the value of the trait may provide information about the disease genotype in unaffected individuals. The analysis of sib-pair data for linkage between a quantitative trait and marker loci is known as QTL (quantitative trait loci) linkage analysis (Risch & Zhang, 1995; Fulker et al, 1995).

Finally, it must be recognised that linkage analysis is inherently limited by the effective size of the disease locus. When the disease allele increases the risk of the disease by a factor less than 2, then the detection of linkage will require enormous samples, whatever method of linkage analysis is used. The same effect may, however, be detectable by allelic association analysis in a realistic sample (Risch & Merigangas, 1996).

7 Conclusion

The affected sib-pair method of linkage analysis is useful primarily for common diseases where the "sibling relative risk" associated with any contributory locus is modest ($1.5 < \lambda_{iS} < 2.5$). Below a "sibling relative risk" of 1.5, the number of affected sib-pairs required to detect linkage becomes

prohibitively large. On the other hand, when the "sibling relative risk" is above 2.5, other types of affected relative-pairs and larger pedigrees become substantially more informative for linkage than affected sib-pairs.

For a sample that consists predominantly of affected sib-pairs (with or without unaffected relatives), the analytical method and computer program of choice is probably MAPMAKER/SIBS. This method is able to make full use of multipoint and incomplete data. If the sample contains an appreciable proportion of families with more complex structure, then an alternative method such as the NPL_{all} test in GENEHUNTER is advisable.

8 References

Blackwelder WC, Elston RC. (1985) A comparison of sib-pair linkage tests for disease susceptibility loci. Genetic Epidemiology, 2, 85-97.

Curtis D, Sham PC. (1994) Using risk calculation to implement an extended relative pair analysis. Annals of Human Genetics, 58, 151-162.

Curtis D, Sham PC. (1995) Model-free linkage analysis using likelihoods. American Journal of Human Genetics, 57, 703-716.

Davis S, Schroeder M, Goldin LR, Weeks DE. (1996) Nonparametric simulation-based statistics for detecting linkage in general pedigrees. American Journal of Human Genetics, 58, 867-880.

Faraway JJ. (1993) Improved sib-pair linkage test for disease susceptibility loci. Genetic Epidemiology, 10, 225-233

Fulker DW, Cherny SS, Cardon LR. (1995) Multipoint interval mapping of quantitative trait loci, using sib-pairs. American Journal of Human Genetics, 56, 1224-1233.

Holmans P. (1993) Asymptotic properties of affected-sib-pair linkage analysis. American Journal of Human Genetics, 52, 362-374.

Knapp M, Seuchter SA, Baur MP. (1994) Linkage analysis in nuclear families. 1. Optimality criteria for affected sib-pair tests. Human Heredity, 44, 37-43.

Kruglyak L, Lander ES. (1995) Complete multipoint sib pair analysis of qualitative and quantitative traits. American Journal of Human Genetics, 57, 439-454.

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander LS. (1996) Parametric and Nonparametric Linkage Analysis: A Unified Multipoint Approach. American Journal of Human Genetics, 58, 1347-1363.

Risch N (1990a) Linkage strategies for genetically complex traits. II.

The power of affected relative pairs. American Journal of Human Genetics, 46, 229-241.

Risch N (1990b) Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs. American Journal of Human Genetics, 46, 242-253.

Risch N, Zhang H. (1995) Extreme discordant sib pairs for mapping quantitative trait loci in humans. Science, 268, 1584-1589

Risch N, Merigangas K. (1996) The future of genetic studies of complex human diseases. Science, 273, 1516-1517.

Sandkuijl LA. (1989) Analysis of affected sib-pairs using information from extended families. Multipoint mapping and linkage based upon affected pedigree members: Genetic Analysis Workshop 6, Alan R Liss, New York.

Self SG, Liang KY. (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. Journal of the American Statistical Association, 82, 605-610.

Suarez BK, Rice JP, Reich T. (1978) The generalised sib-pair IBD distribution: Its use in the detection of linkage. Annals of Human Genetics, 42, 87-94.

Suarez BK, Van Eerdewegh P. (1984) A comparison of three affected-sib-pair scoring methods to detect HLA-linked disease susceptibility genes. American Journal of Medical Genetics, 18, 135-146.

9 List of Computer programs

```
ANALYZE
Joe Terwilliger
ftp://ftp.well.ox.ac.uk/pub/genetics/analyze or
ftp://linkage.cpmc.columbia.edu/software/analyze

APM
Daniel E. Weeks (dweeks@watson.hgen.pitt.edu),
Mark Schroeder (mark@holmes.hgen.pitt.edu)
ftp://watson.hgen.pitt.edu/pub/apm

ASPEX
David Hinds ( dhinds@lahmed.stanford.edu )
Neil Risch
http://lahmed.stanford.edu/pub/aspex/index.html
```

ftp://lahmed.stanford.edu/pub/aspex

ERPA
Dave Curtis (dcurtis@hgmp.mrc.ac.uk)
http://www.gene.ucl.ac.uk/~dcurtis/software.html
ftp://ftp.gene.ucl.ac.uk/pub/packages/dcurtis

ESPA
Lodewijk Sandkuijl (sandkuyl@rullf2.leidenuniv.nl)

GAS
Alan Young (ayoung@vax.ox.ac.uk)
http://users.ox.ac.uk/~ayoung/gas.html
http://linkage.rockefeller.edu/soft/gas/gas2.html
ftp://ftp.ox.ac.uk/pub/users/ayoung
ftp://linkage.rockefeller.edu/software/gas

MAPMAKER/SIBS
Leonid Kruglyak (leonid@genome.wi.mit.edu),
Mark Daly (mjdaly@genome.wi.mit.edu),
Eric Lander (lander@genome.wi.mit.edu)
ftp://ftp-genome.wi.mit.edu/distribution/software/sibs

SAGE
R. Elston, J. Bailey-Wilson, G. Bonney, L. Tran, B. Keats,A. Wilson
http://darwin.mhmc.cwru.edu/ (previously:
http://csmith.biogen.lsumc.edu)

SIMIBD
Sean Davis (davis@moriarty.hgen.pitt.edu) and Daniel E. Weeks
ftp://watson.hgen.pitt.edu/pub/simibd

SPLINK
David Clayton
ftp://ftp.mrc-bsu.cam.ac.uk/pub/methodology/genetics

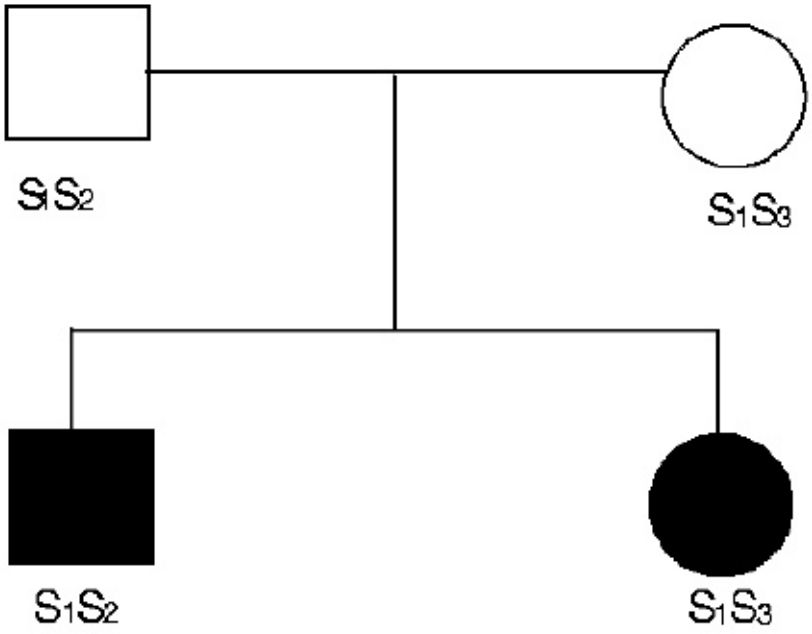

Figure 1: Illustrative family

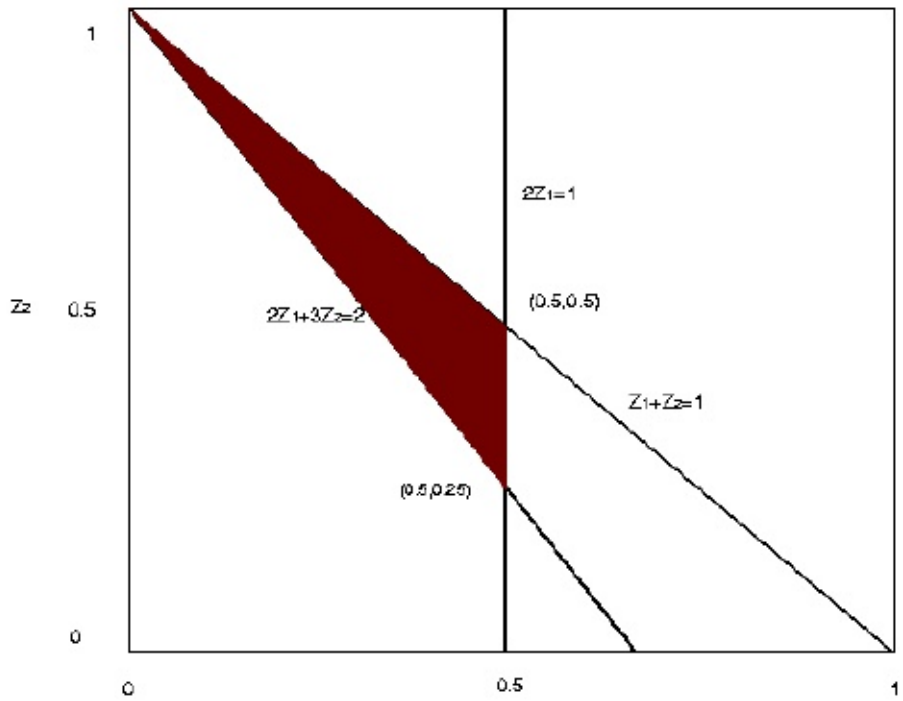

Figure 2: The possible triangle for IBD probabilities